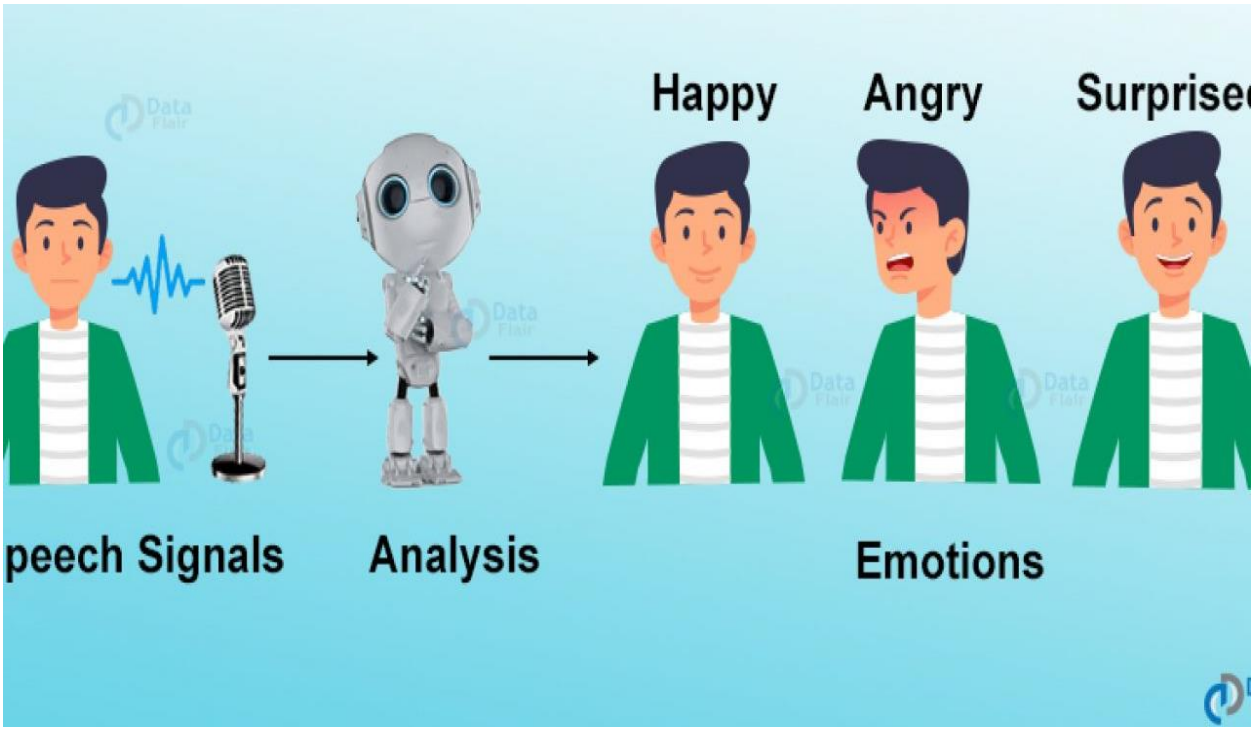
SPEECH EMOTION RECOGNITION



What is Speech Emotion Recognition?

Speech Emotion Recognition, abbreviated as SER, is the act of attempting to recognize human emotion and affective states from speech. This is capitalizing on the fact that voice often reflects underlying emotion through tone and pitch. This is also the phenomenon that animals like dogs and horses employ to be able to understand human emotion.

SER is tough because emotions are subjective and annotating audio is challenging.

What is librosa?

librosa is a Python library for analyzing audio and music. It has a flatter package layout, standardizes interfaces and names, backwards compatibility, modular functions, and readable code. Further, in this Python mini-project, we demonstrate how to install it (and a few other packages) with pip.

Speech Emotion Recognition – Objective

To build a model to recognize emotion from speech using the librosa and sklearn libraries and the RAVDESS dataset.

Speech Emotion Recognition – About the Project

In this project, we will use the libraries librosa, soundfile, and sklearn (among others), keras to build a model. This will be able to recognize emotion from sound files. We will load the data, extract features from it, then split the dataset into training and testing sets. Then, we'll initialize an algorithm and train the model. Finally, we'll calculate the accuracy of our model.

The Dataset

For this project, you are provided with the RAVDESS dataset; this is the Ryerson Audio-Visual Database of Emotional Speech and Song dataset. This dataset has 7356 files rated by 247 individuals 10 times on emotional validity, intensity, and genuineness. The entire dataset is 24.8GB from 24 actors, but we've lowered the sample rate on all the files to reduce the memory to few MBs.

Prerequisites

You'll need to install the following libraries with pip:

```
pip install librosa soundfile numpy sklearn pyaudio
```